

# LearnOSM

## Revisione di Dati OSM

Questa guida può essere scaricata come [review\\_it.odt](#) o [review\\_it.pdf](#)  
Rivisto il 2017-09-15

- TOC { :toc }

**Questa sezione descrive i processi per il controllo della qualità dei dati, in particolare nel contesto di un progetto di mappatura OSM, come quelli intrapresi da [Humanitarian OpenStreetMap Team](#) in vari paesi ed i progetti [Open Cities](#) in Bangladesh, Sri Lanka, e Nepal. Le metodologie descritte possono essere utili anche in altri contesti, in cui sia previsto il controllo della qualità dei dati.**

Quando cerchiamo di mappare un insieme completo di caratteristiche e attributi in una specifica area, abbiamo bisogno di metodi per il controllo degli errori e di metodi per valutare l'accuratezza del lavoro. In questo tutorial useremo diversi metodi per il controllo dei dati, spiegando i passi del processo e la loro motivazione. Un progetto di mappatura ben gestito includerà questi tre processi: sia la valutazione e correzione dei dati, sia la loro rendicontazione.

- Controlli giornalieri
- Riesame
- Query SQL

Questi metodi di revisione diventano più importanti al crescere del modello dei dati e all'aumentare del numero di caratteristiche rilevate. Per esempio, non richiederebbe molto tempo e sforzo valutare un modello dati che riguarda solo dei punti di interesse (POI):

Data Model	
POIs	Name
	Type
	Phone Number

In questo caso le domande a cui rispondere sarebbero:

- I POI sono stati mappati in tutte i luoghi?
- Ci sono POI senza l'attributo del nome?
- Ci sono POI senza l'attributo del tipo?
- Ci sono POI senza l'attributo del numero telefonico?
- Sono indicate correttamente le lettere maiuscole nel nome?
- Il numero telefonico è valido?

Di solito un modello dati è molto più complesso, come nel caso della mappatura di edifici. Consideriamo un modello dati che include questo:

Buildings	Name
	Street Name
	Building Number
	Number of Levels
	Use of Building
	Material of Building
	Type of Structure
	Use of Building
	Construction Date
	Building Condition

In questo momento potreste mappare migliaia di edifici che hanno molti attributi e l'analisi diventa più critica. In questo tutorial useremo gli edifici come esempio, ma gli stessi metodi possono applicarsi anche per la revisione di altri tipi di elementi.

## Controlli giornalieri

Il modo più diretto di controllare i dati è di esaminarli e validarli regolarmente. Questo può essere fatto giornalmente o al massimo settimanalmente. Per il supervisore di una squadra di mappatori, questa è un'attività importante perché rilevare il prima possibile errori e modalità sbagliate di editare significa poterli correggere e i mappatori possono imparare a fare l'attività nel modo giusto.

Ora esamineremo alcuni metodi per controllare i dati semplicemente usando JOSM. Alcune delle domande che ci facciamo sono:

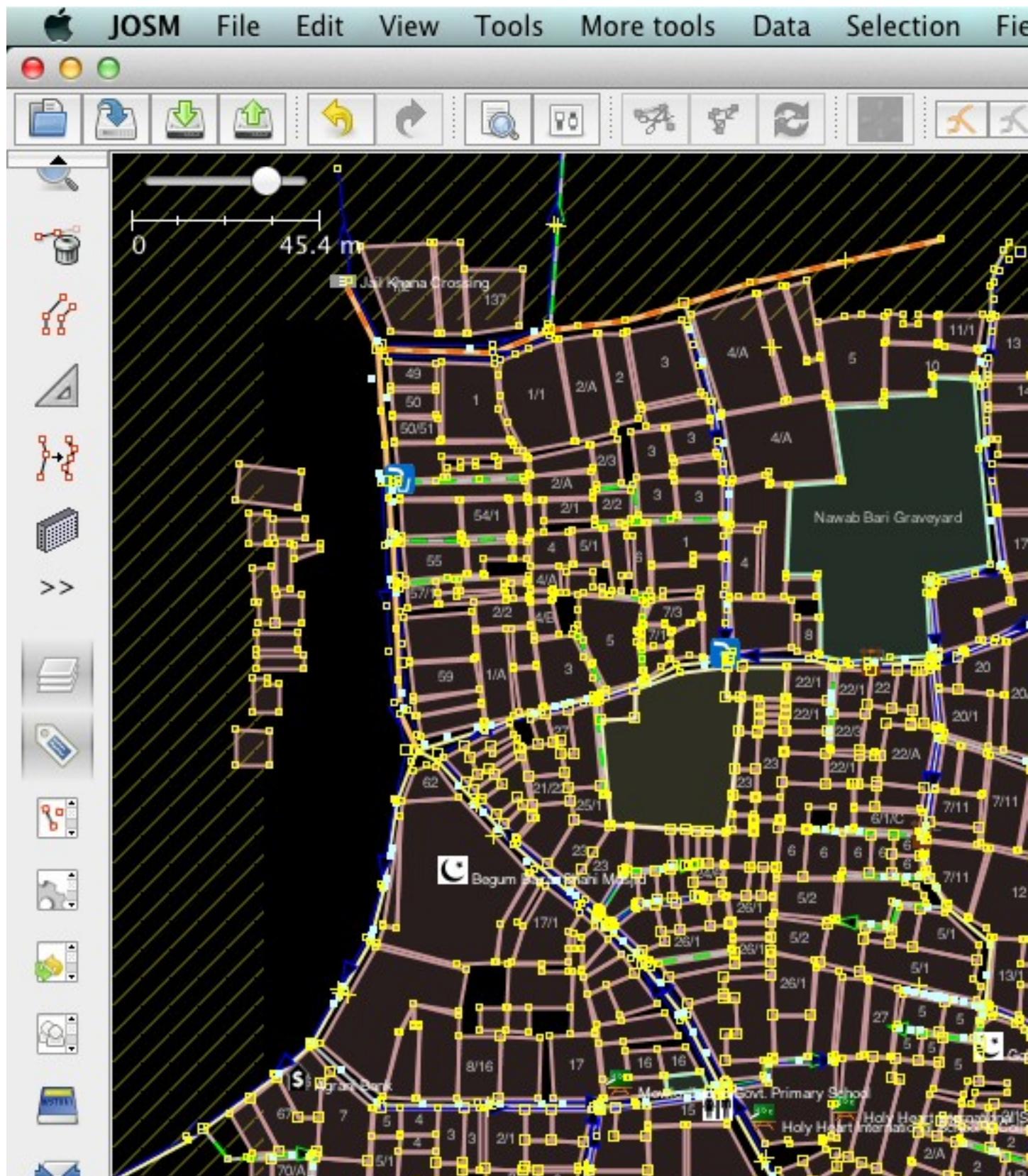
- Ci sono errori di **topologia** (come la sovrapposizione di edifici o relazioni sbagliate)?
- Ci sono errori di **etichettatura** (etichette con errori di battitura, combinazioni di chiave-valore improprie)?
- i dati sono **completi** rispetto al modello dei dati?

Esaminiamo come possiamo rispondere a queste domande con JOSM. Stiamo ipotizzando di esaminare il lavoro di altri, ma gli stessi processi funzionano bene (e dovrebbero essere più facili) quando esaminiamo il nostro stesso lavoro.

Useremo come esempio un file dal progetto di mappatura di Open Cities a Dhaka. Per seguirlo, scaricate questo file: [dhaka\\_validation\\_example.osm](http://dhaka_validation_example.osm)

([http://learnosm.org/files/dhaka\\_validation\\_example.osm](http://learnosm.org/files/dhaka_validation_example.osm))

**NON cercate di salvare le vostre modifiche in OpenStreetMap. Questi esercizi hanno solo scopo dimostrativo.**



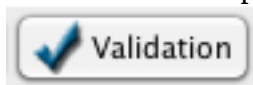
## Validazione dei dati

Il primo passo per controllare i dati consiste nell'eseguire lo strumento di Validazione in JOSM, che controllerà i dati per errori sospetti. Questo strumento è particolarmente utile per trovare errori di **topologia** ma può non essere altrettanto utile per trovare etichette errate.

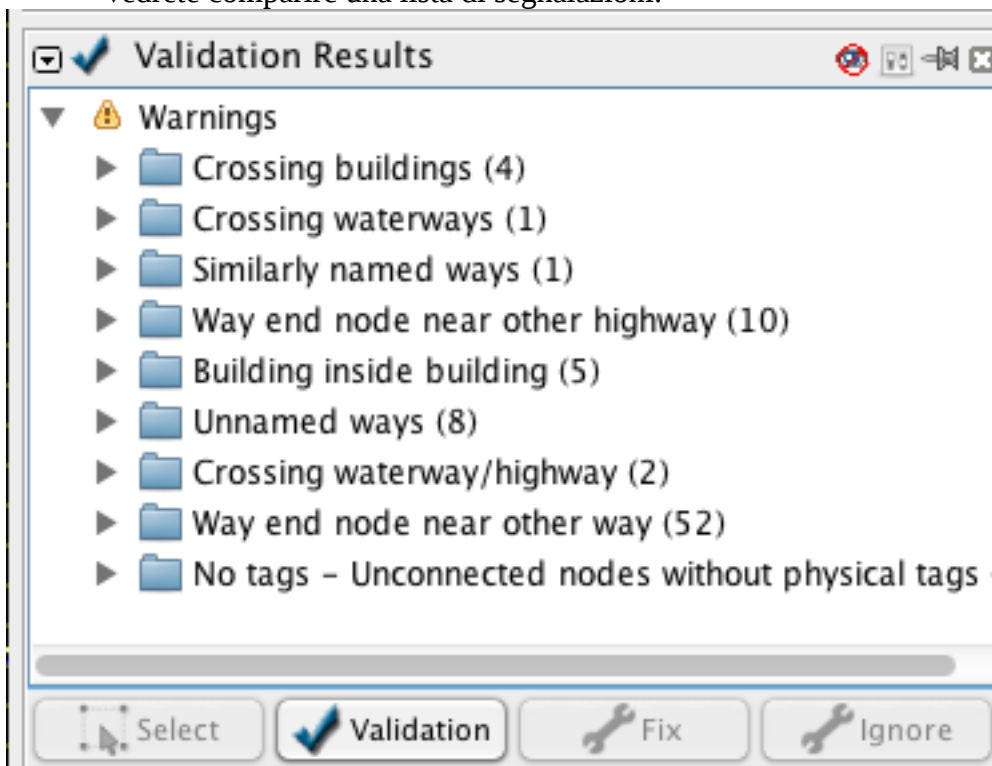
- Attivate lo strumento cliccando il pulsante dello strumento di Verifica nel lato sinistro di JOSM. (Questo non è necessario se il pannello di Verifica è già aperto)



- Poi assicuratevi che non sia selezionato niente, cliccando su un punto vuoto della mappa. Se avete elementi selezionati quando eseguite lo strumento di Verifica, saranno controllati solo quegli elementi. (A volte potreste voler controllare solo certi elementi, ma per ora controlleremo l'intero file)
- Cliccate il pulsante "Convalida" nel pannello.

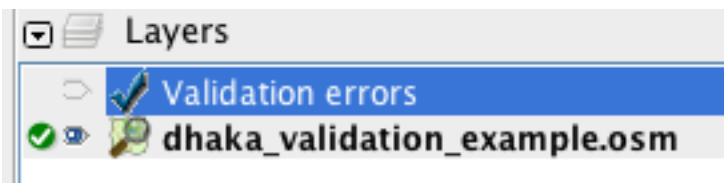


- Vedrete comparire una lista di segnalazioni:

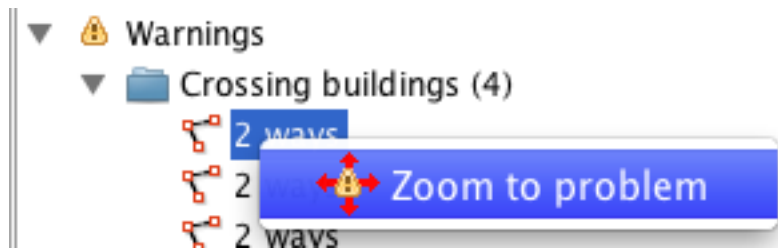


- Comparirà anche un altro Livello, che mostra dove sono gli errori. Può esservi utile nascondere questo livello, per ora.

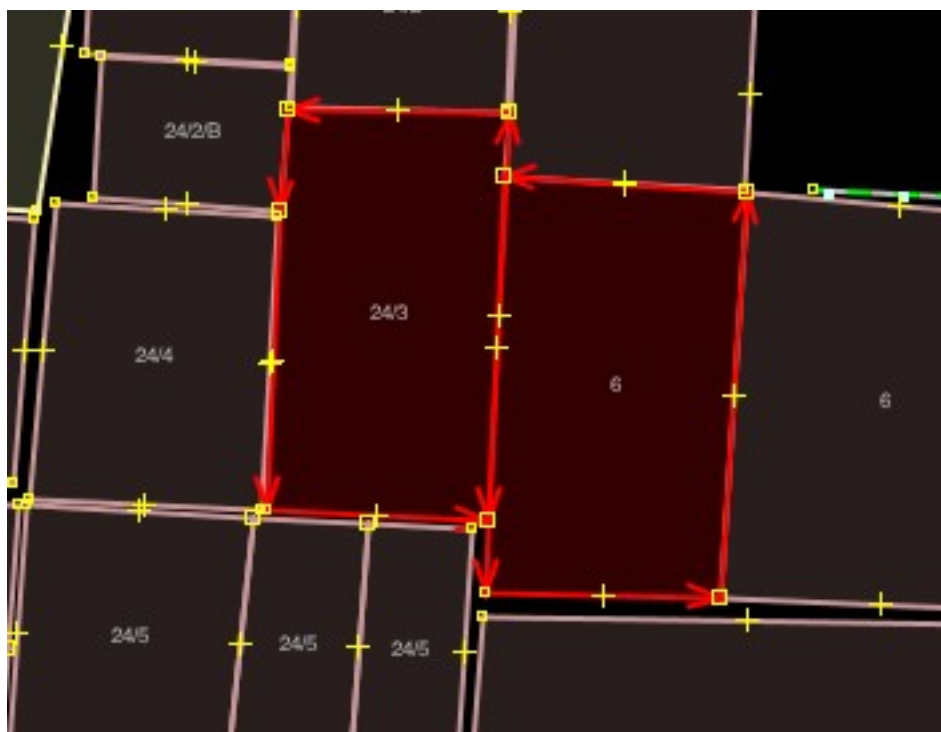




Diamo un'occhiata ad alcune segnalazioni. Potete vedere che ci sono quattro segnalazioni "edifici sovrapposti". Questa segnalazione significa che degli edifici in qualche modo si sovrappongono. Selezionate il primo elemento di questa lista, fate click con il pulsante destro del mouse, e cliccate "Ingrandisci visuale su problema."



Cliccate anche il pulsante "Seleziona" in fondo alla finestra di Verifica per selezionare le linee in questione. Viene così mostrato che queste due linee hanno un problema:

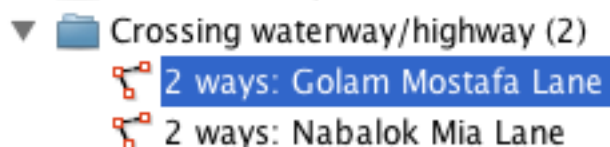


- Questo è un errore che non avremmo mai notato senza lo strumento di verifica. Se zoomate parecchio vedrete che c'è una piccola sovrapposizione tra i due edifici, che è un errore topologico, perché gli edifici solitamente non si sovrappongono uno all'altro. Per risolverlo, il nodo intermedio deve essere spostato. Se gli edifici sono a contatto, cosa probabile, il nodo intermedio può essere aggiunto alla linea.

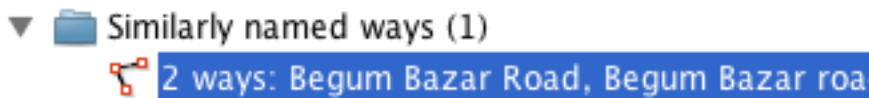
- Una volta eseguita questa correzione, possiamo nuovamente eseguire lo strumento di Verifica e questa segnalazione non comparirà più nell'elenco.

Questo metodo di controllare automaticamente i dati è un modo efficace di correggere errori topologici, in particolare quelli che sarebbe difficile notare di persona. Nell'elenco delle segnalazioni di verifica potete vedere che altre segnalazioni, come "Edificio interno ad altro edificio", sono conseguenza di errori simili.

A volte delle segnalazioni, come "Corsi d'acqua/strade incrociate", non sono necessariamente degli errori. Questo indica che lo strumento di verifica è in grado di trovare possibili errori, ma richiede che qualcuno vada e controlli se l'errore esiste o no.



Guardiamo la segnalazione sotto "Percorsi con nomi simili" per vedere un errore che non è topologico. Cliccate "Seleziona" per selezionare le due linee in questione.

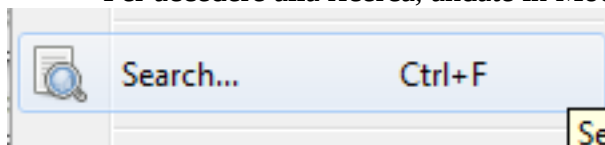


Sapete dire di che errore si tratta? In questo caso abbiamo due segmenti diversi di strada, che sono effettivamente la stessa strada, tuttavia hanno dei nomi leggermente diversi - "road" inizia in maiuscolo in una delle strade, ma non nell'altra. E' probabile che debbano avere lo stesso nome, e in questo caso la parola "road" dovrebbe iniziare in maiuscolo.

## Utilizzo della funzione di ricerca in JOSM

La ricerca in JOSM è un modo efficace di riesaminare i dati. Consente di indicare dei termini di ricerca, noti anche come query, per selezionare solo gli elementi desiderati.

- Per accedere alla ricerca, andate in Modifica -> Cerca o premete CTRL + F da tastiera.



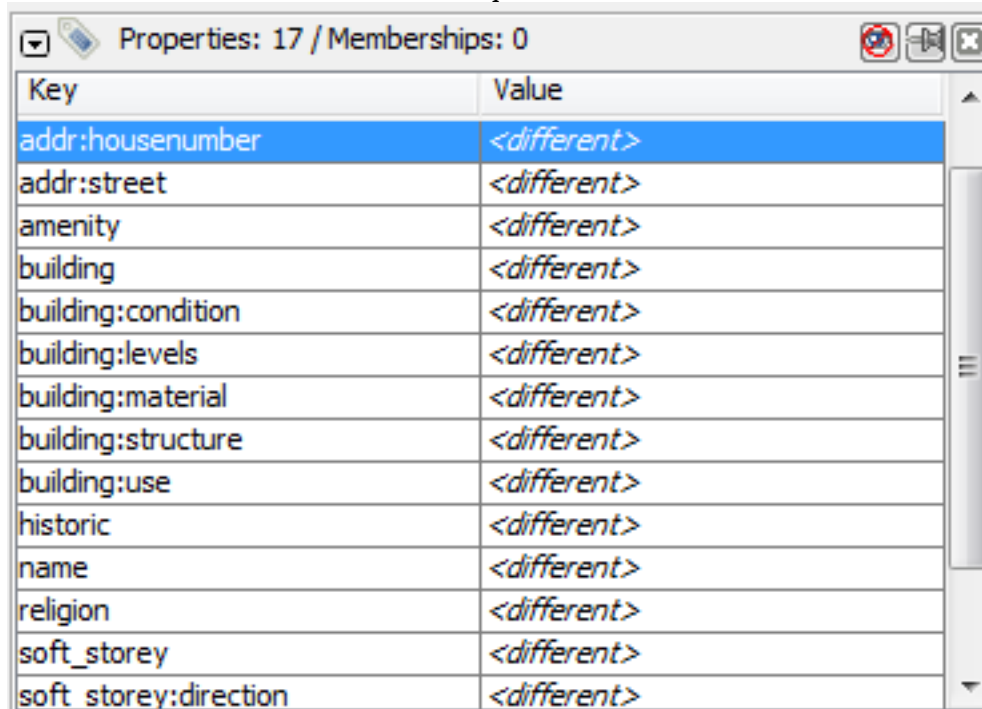
- Ci sono un bel po' di tipi di interrogazioni che qui potete effettuare, e potete vedere dettagli ed esempi nella stessa finestra di ricerca e cliccando sul pulsante "Aiuto".
- Per ora, selezionate tutti gli edifici. Quasi tutti gli edifici hanno l'etichetta **building=yes** e alcuni hanno **building=construction**. Questo indica che possiamo costruire una query con:  
*building = yes OR building=construction*
- Così selezioniamo tutti gli edifici, ma nel caso qualcuno abbia associato un'etichetta sbagliata ad un edificio, possiamo usare invece un carattere jolly, che selezionerà tutti gli elementi che hanno come chiave **building**.

Search string: `building = *`

- Tutti gli edifici saranno selezionati.

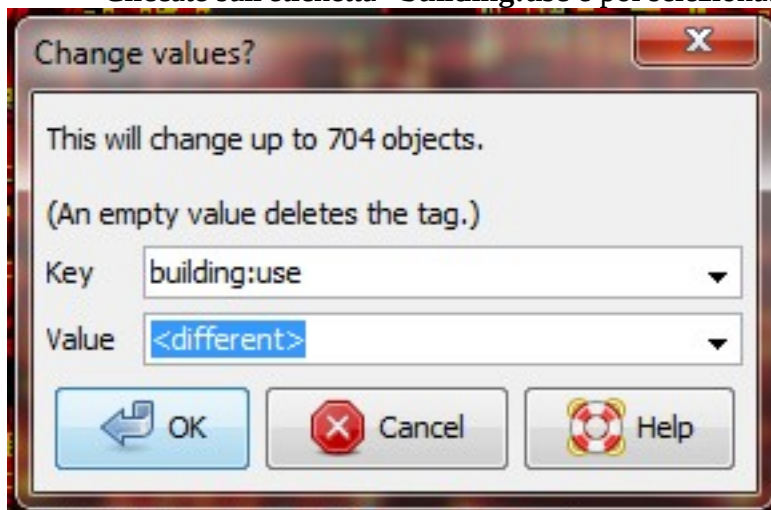
Questo è ottimo, ma come ci aiuta ad esaminare i dati? Beh, ora che è stato selezionato tutto quello relativo ad un singolo elemento, possiamo cercare etichette errate.

- Guardate nella finestra delle Etichette - quello che vediamo sono tutte le etichette di tutti gli oggetti selezionati. Questi condividono tutti la stessa chiave, ma siccome ogni elemento ha valori differenti, questi sono marcati come .



Key	Value
addr:housenumber	<different>
addr:street	<different>
amenity	<different>
building	<different>
building:condition	<different>
building:levels	<different>
building:material	<different>
building:structure	<different>
building:use	<different>
historic	<different>
name	<different>
religion	<different>
soft_storey	<different>
soft_storey:direction	<different>

- Cliccate sull'etichetta **\*building:use** e poi selezionate "Modifica".



Change values?

This will change up to 704 objects.  
(An empty value deletes the tag.)

Key: `building:use`

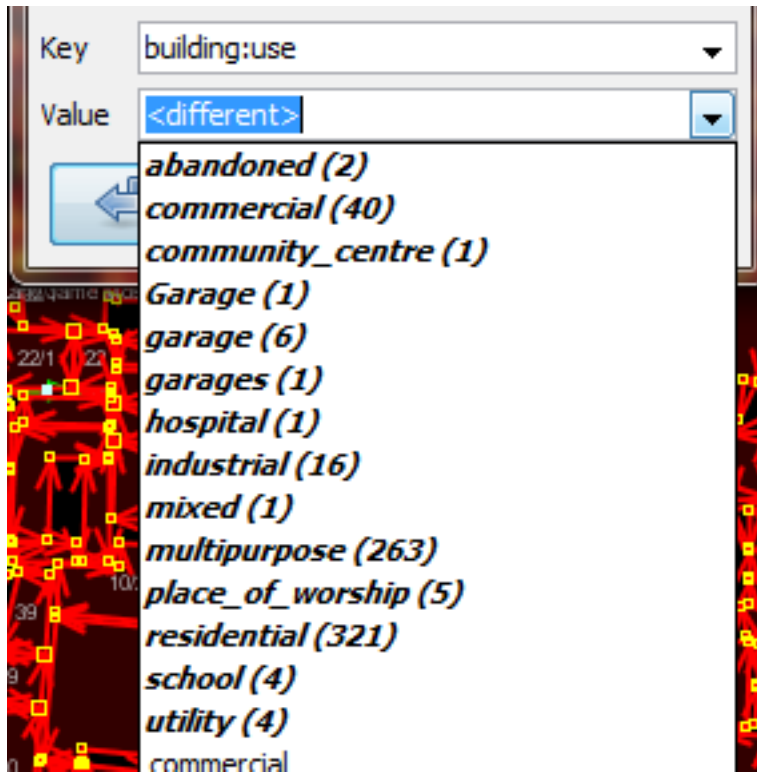
Value: `<different>`

Buttons: OK, Cancel, Help

- QUI FATE ATTENZIONE! Non modificate il valore e cliccate OK, perché questo

cambiarebbe il valore dell'etichetta per tutti gli oggetti che hanno la caratteristica 'building'. **Questo sarebbe del tutto sbagliato.**

- Cliccate invece sulla casella a discesa vicino a Valore.



- Notate che tutti gli elementi in grassetto hanno vicino un numero tra parentesi. Questo è il numero di elementi selezionati che hanno questo valore di etichetta.

Possiamo confrontare questo con le etichette OpenStreetMap che sono state mappate nel nostro modello dati, e cercare errori. Per esempio, questa etichetta rappresenta l'utilizzo di un edificio. All'inizio nel progetto Open Cities Dhaka (da cui provengono questi dati) c'era incertezza se etichettare un edificio con uso misto con ***building:use=multipurpose*** oppure ***building:use=mixed***. Visto che la prima etichetta era stata usata precedentemente in altre nazioni, fu selezionata questa. Tuttavia, vediamo che uno degli edifici è stato etichettato con ***mixed***. Dobbiamo correggerlo (Un altro ovvio errore sono i tre valori diversi per ***garage***, ma questi li tralasciamo ora).

- Non possiamo cambiare adesso l'elemento che ha l'etichetta ***building:use=mixed***, perché abbiamo centinaia di oggetti selezionati. Perciò, per correggere l'errore, dobbiamo prima trovarlo. Come? Avete indovinato - con lo strumento Cerca.
- Cliccate "Annulla" per uscire da questa finestra. **Ricordate, cliccare OK può essere pericoloso.**
- Aprite di nuovo la finestra Cerca ed inserite la seguente query: ***"building:use"=mixed***
- Notate che le virgolette sono necessarie perché i due punti (:) hanno un loro significato di ricerca. Questo farà sì che sia selezionato il solo edificio che ha questa etichetta, che ora può essere corretto con il valore ***multipurpose***.

**Ricordate che mentre seguite questo tutorial, NON provate a salvare le vostre modifiche in**



**OpenStreetMap. Questi esercizi sono solo a scopo dimostrativo.**

## ***Riesame***

Quando si gestisce un progetto come il rilevamento dettagliato dell'edificio, ci dovrebbe essere un metodo ulteriore di controllo qualità, sia per migliorare il lavoro sia per documentarne l'accuratezza alla fine del progetto.

Se ci sono molte squadre di mappatori che cooperano nel rilevamento di un'area, è probabile che una o più squadre non facciano un lavoro soddisfacente. Anche quelle squadre che fanno un lavoro efficiente ed accurato faranno degli errori. Pensate a delle squadre che mappano ciascuna 100 edifici al giorno - non è improbabile che una piccola percentuale degli attributi che rilevano sia sbagliata.

Così, un buon progetto includerà un processo di ricontrollo di parte del lavoro che è stato fatto, correzione di errori, determinare quali squadre di mappatori stanno operando in modo soddisfacente, e approssimare la percentuale di errori per la rendicontazione finale.

Certo, non ha senso ricontrollare ogni edificio in una data area, ma il 5-10% degli edifici dovrebbe essere rivisto. Le zone da revisionare dovrebbero essere scelte in aree differenti per poter confrontare le squadre di rilevazione. Le squadre di rilevazione possono ricontrollarsi tra di loro il lavoro fatto, o se possibile gestori più esperti possono occuparsi dei controlli. E' una pratica diffusa che i gestori dedichino un giorno alla settimana riesaminando parti dell'area rilevata.

## **Correzione errori**

Cosa si dovrebbe fare quando si trovano errori?

Se c'è un piccolo numero di errori (meno del 5% degli edifici), i problemi dovrebbero essere riportati alla squadra originaria di rilevamento, così che ne sia informata e non possa ripetere gli stessi errori nuovamente. I dati dovrebbero essere corretti in OpenStreetMap ed i risultati del nuovo rilevamento dovrebbero essere registrati.

Se ci sono molti errori, può essere necessario intraprendere azioni più significative. La squadra di rilevamento avrà bisogno di essere affrontata in modo appropriato, e le aree da lei mappate possono aver bisogno di essere completamente rilevate da capo, a seconda di quanto i dati risultino inesatti. Una inaccuratezza superiore al 10% molto probabilmente è una percentuale non accettabile.

## **Rendicontazione dell'Accuratezza**

Il secondo obiettivo del ricontrollo è quello di poter rendicontare l'accuratezza dei dati quando il progetto termina. Gli utilizzatori dei dati vorranno sapere quali metriche e metodologie per stimare la qualità dei dati sono state usate.

Includendo questo processo nella vostra metodologia di revisione, sarete in grado di spiegare chiaramente come avete stimato la qualità dei dati, e fornire i numeri che mostrano la percentuale probabile di errore contenuta nei dati che avete esaminato.

Per esempio, immaginiamo di stare gestendo un progetto che mappa 1000 edifici. Decidiamo quindi di mapparne il 10%, o 100 edifici, selezionati casualmente nell'area di rilevamento. Andiamo sul campo e scopriamo che dei 100 edifici da ricontrollare, sei di questi hanno un alto livello di inaccuratezza. Diciamo che consideriamo un rilevamento inaccurato se ha più di un attributo sbagliato. Dunque il 6% di quanto controllato è errato - possiamo correggere questi errori, ma dobbiamo comunque estrapolare che circa il 6% dei 1000 edifici sono probabilmente inaccurati. Questo dovrebbe essere riportato come il probabile errore alla chiusura del progetto.

Il riesame dovrebbe essere condotto durante tutto il progetto. Pensate se aspettassimo la fine del progetto, in questo esempio, e fossero sbagliati il 40 per cento degli edifici! Potrebbe rovinare l'intero progetto. E' meglio rilevare errori su larga scala quanto prima, in modo da poterli correggere.

## Query SQL

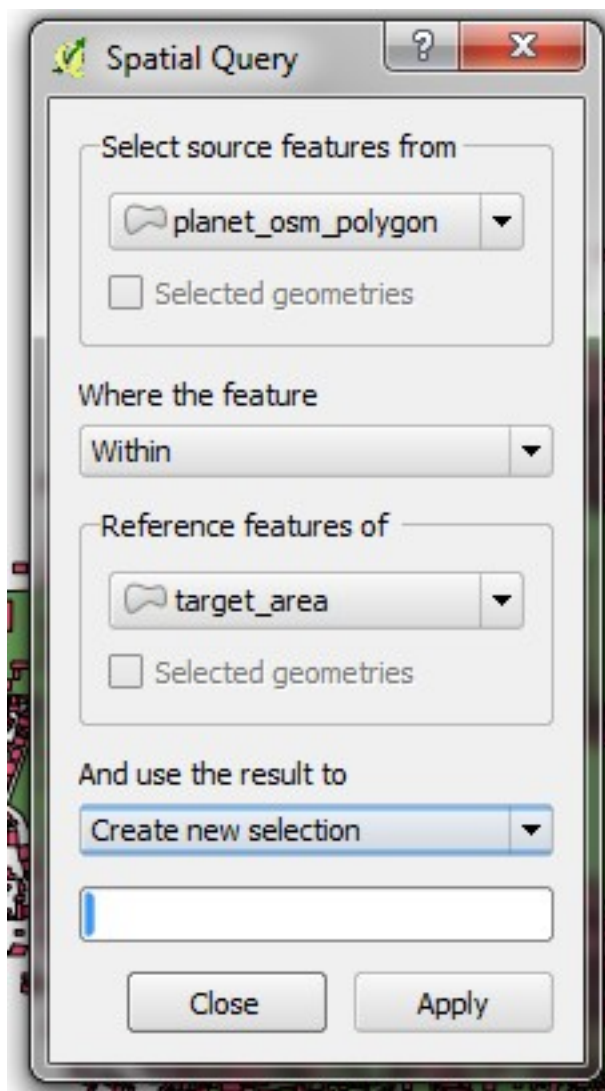
Probabilmente il migliore strumento di analisi è quello di eseguire query SQL in un sistema GIS, tipo Quantum GIS. Questo è simile a cercare i dati in JOSM, ma consente analisi più potenti, anche se può richiedere un po' più di tempo la preparazione. Usare JOSM è un modo veloce, standard per controllare gli errori basilari, mentre fare interrogazioni in QGIS è più efficace per trovare dati mancanti o attributi sbagliati.

Ipotizziamo che siate in qualche modo familiari con i GIS, e ci concentriamo sulla costruzione di query che possono aiutarvi nella revisione dei dati OpenStreetMap. Per gli esercizi che seguono useremo ancora i dati dal progetto Open Cities Dhaka, che potete scaricare da [dhaka\\_sql.zip](http://dhaka_sql.zip). I dati OpenStreetMap sono stati esportati usando lo strumento HOT Export Tool ([export.hotosm.org](http://export.hotosm.org)) ed il perimetro dell'area da rilevare era stata definita all'inizio del progetto.

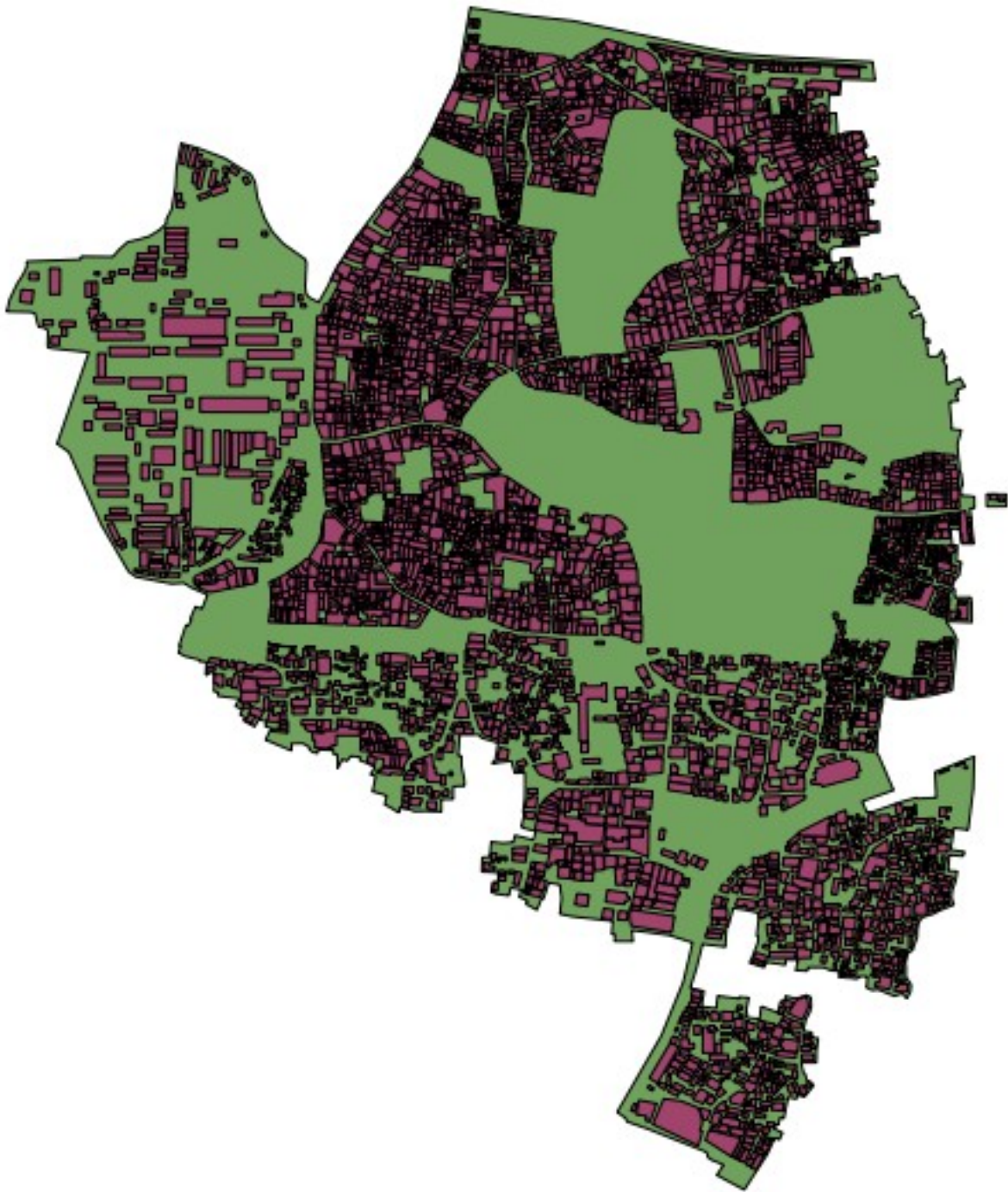
## Preparazione dei dati

Decomprimete i file e caricate i due shapefile in QGIS. Cominceremo prendendo solo gli edifici dell'area del progetto, per rendere le nostre query più semplici.

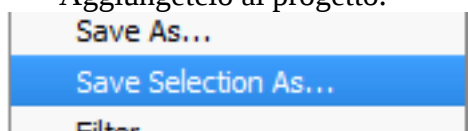
- Per prima cosa selezioniamo solo i poligoni che sono dentro l'area esaminata. Per fare questo useremo il plugin Spatial Query ("Plugin di interrogazione spaziale"). Se non l'avete già installato, andate a Plugins -> Gestisci e Installa Plugin per trovarlo ed installarlo.
- Andate a Vettore -> Interrogazione Spaziale -> Interrogazione Spaziale
- Impostate come parametri per selezionare gli elementi da **planet\_osm\_polygon** che sono **contenuto** dentro **target\_area**.



- Cliccate Apply. Solo i poligoni dentro l'area da esaminare saranno selezionati.



- Cliccate con il pulsante destro sul livello e salvate la selezione come nuovo shapefile. Aggiungetelo al progetto.

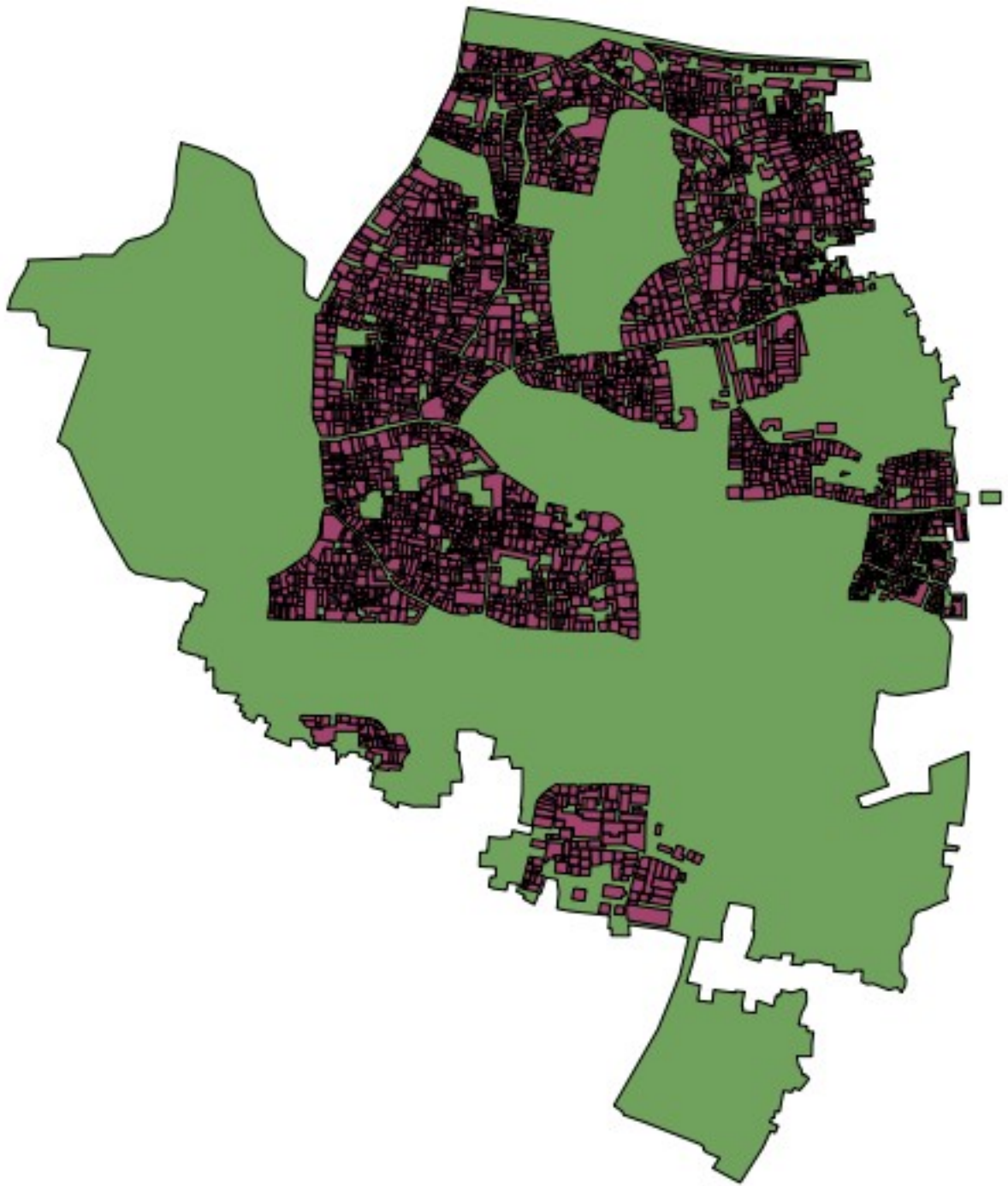


- Ora filtriamo solo i poligoni che sono edifici e che sono stati rilevati nell'ambito di questo progetto.
- Clic con il pulsante destro su **planet\_osm\_polygon** e poi cliccate su "Filtro..."
- Inserite la seguente query:

*"building" != NULL AND "source" = 'Open Cities Dhaka Survey'*

- Cliccate OK. Filtrando i dati con questa query saranno mostrati solamente i poligoni che hanno qualche valore nella colonna 'building'. Saranno anche rimossi gli edifici che non hanno un'etichetta 'source' associata con questo progetto.
- Salvate questi dati come nuovo shapefile. Useremo questo file per le nostre query SQL.





## Query SQL

Ora possiamo eseguire query sul livello degli edifici per trovare possibili errori. Pensiamo a qualcosa che potremmo voler interrogare. Il modello dati di questo progetto indica gli attributi

che dovrebbero essere raccolti per ogni edificio - questi sono:

- name (nome)
- building (edificio)
- building:levels (numero di piani)
- building:use (utilizzo)
- building:vertical\_irregularity (irregolarità verticali)
- building:soft\_storey (edifici con debolezze)
- building:material (materiale di costruzione)
- building:structure (struttura)
- start\_date (data di costruzione)
- building:condition (condizioni)

Notate che nello shapefile questi nomi di attributi sono troncati, dato che i nomi delle colonne sono limitati a 10 caratteri.

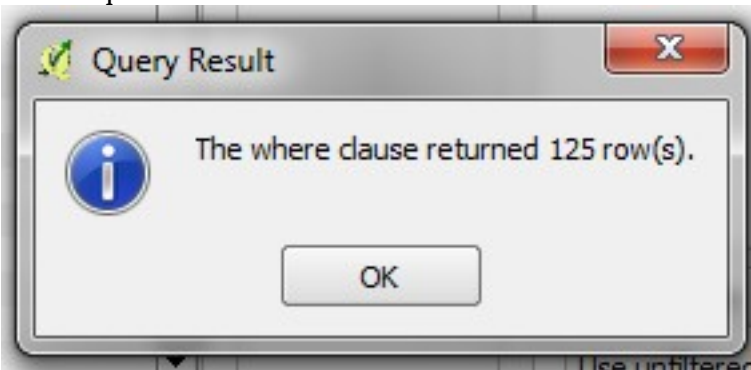
Dunque che tipo di domande vogliamo fare? Quali sono i probabili errori? Un errore comune è quello che un edificio è stato mappato, ma non tutti i suoi attributi sono stati raccolti. Allora potremmo fare una query che mostra tutti gli edifici che non hanno un insieme completo di attributi. Certo, per qualche attributo, come il nome e la data di inizio (l'anno di costruzione), può andare bene che siano vuoti, perchè non tutti gli edifici hanno un nome e a volte l'anno di costruzione non è noto. Ma gli altri attributi dovrebbero sempre essere raccolti.

Proviamo a scrivere una query per questo:

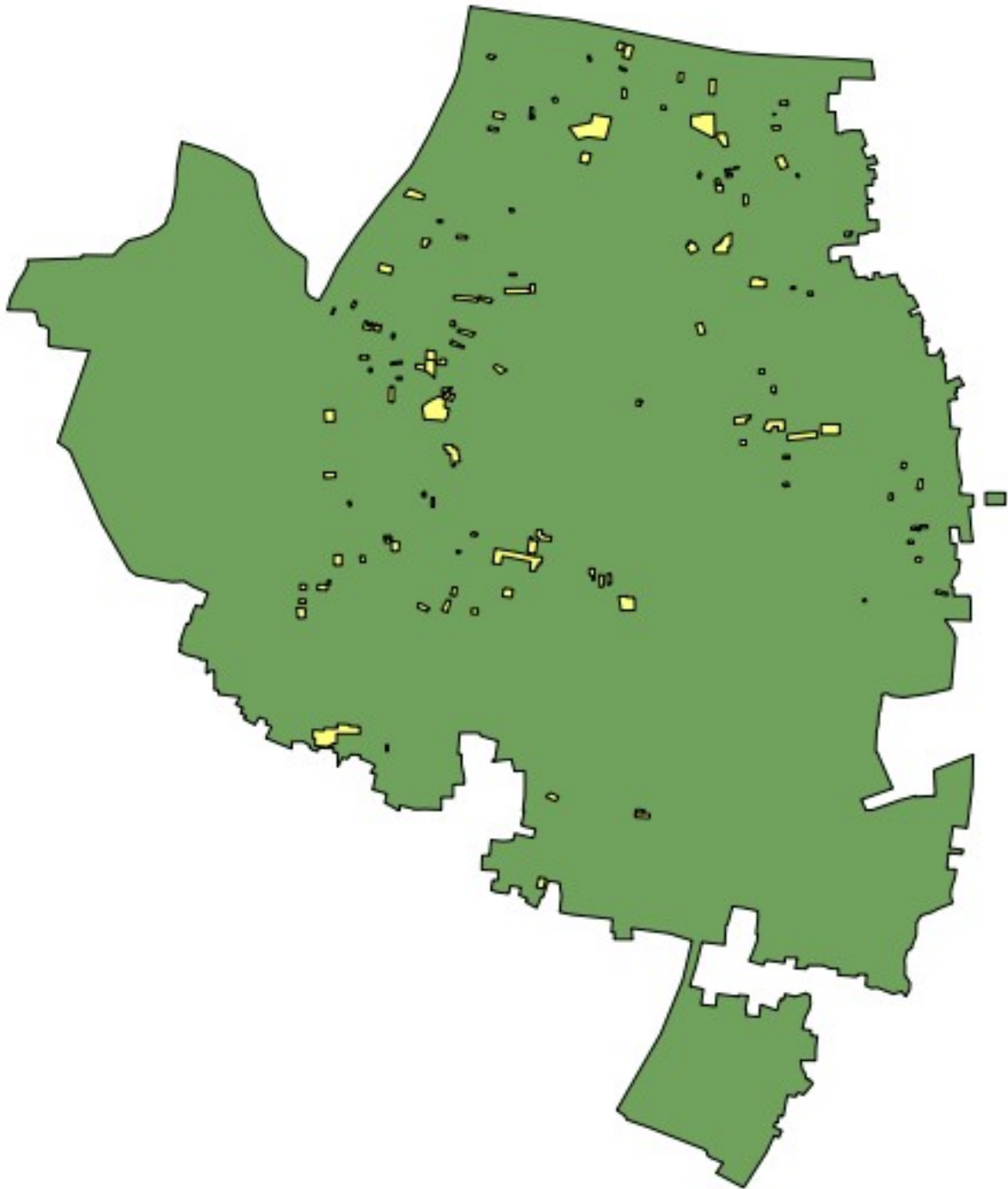
- Clic con il pulsante destro sul livello degli edifici (il livello che abbiamo creato nella sezione precedente) e clic su "Filtro...". Questo aprirà la finestra per il costruttore di query. Qui si possono scrivere query complesse che restituiscono solo i dati che vogliamo.
- Potete costruire la vostra query facendo doppio clic sui campi, gli operatori, e i valori, o potete copiare la query che abbiamo qui costruito:

*"building\_c" = NULL OR "building\_s" = NULL OR "building\_l" = NULL OR "building\_m" = NULL OR "vertical\_i" = NULL OR "soft\_store" = NULL OR "building\_u" = NULL*

- Cliccate "Test" e vedrete che questa query restituisce 125 elementi. Questo significa che dei circa 3500 edifici che sono stati mappati, 125 di questi sono privi di uno o più di questi attributi.



- Cliccate OK per mostrare solo quegli edifici che soddisfano le condizioni della query.



- Questi edifici possono essere esaminati più in dettaglio per identificare quali attributi sono mancanti, e se è necessario che vengano rimappati. E' quindi possibile usare QGIS per creare una mappa che una squadra di ricontrollo può utilizzare per correggere gli attributi degli edifici mancanti.

Quali altre query potrebbero essere utili? Beh, potreste volere controllare anche gli attributi che

non sono previsti nel vostro schema dati. L'abbiamo fatto nella sezione di ricerca in JOSM. Potete costruire una query per trovare tutti gli edifici i cui attributi non corrispondono al vostro modello dati.

Potete anche usarlo per ricercare anomalie, che probabilmente ma non necessariamente sono errori. Per esempio, se apriamo il costruttore di query, selezioniamo **building\_1**, e clicchiamo "Tutto" sotto il riquadro dei Valori per caricare tutti i possibili valori, vediamo che la maggioranza degli edifici ha un numero compreso tra 1 e 20 (questo attributo è building:levels, il numero di piani dell'edificio). Ma c'è anche un valore di 51. Sembra improbabile che ci sia un edificio di 51 piani che sovrasta tutto in quest'area, così possiamo localizzarlo e fare una nota per i mappatori che lo controllino.

L'uso delle query può essere un modo efficace per cercare possibili errori nell'insieme dei dati. In aggiunta ad altre caratteristiche di QGIS, può essere usato per produrre mappe che possono essere usate per riesaminare i dati in un'area.

## **Sommario**

In questo tutorial abbiamo esaminato diversi metodi efficaci per mantenere la qualità dei dati durante un progetto e abbiamo fatto degli esercizi pratici per familiarizzare con la revisione di dati OSM. Quando si organizza un progetto di mappatura, o anche quando si controllano i dati in un'area per uso personale, questi metodi possono tornare utili.